A multispecies approach reveals hot spots and cold spots of diversity and connectivity in invertebrate species with contrasting dispersal modes
Abigail E. Cahill, Aurelien de Jode, Sophie Dubois, Zoheir Bouzaza, D. Aurelle, Emilie Boissin, Olivier Chabrol, Romain David, Emilie Egea, Jean-Baptiste Ledoux, et al.

To cite this version:

HAL Id: hal-01681650
https://hal.archives-ouvertes.fr/hal-01681650
Submitted on 29 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A multispecies approach reveals hot spots and cold spots of diversity and connectivity in invertebrate species with contrasting dispersal modes

Abigail E. Cahill1,2 | Aurélien De Jode1 | Sophie Dubois1 | Zoheir Bouzaza3 | Didier Aurelle1 | Emilie Boissin4 | Olivier Chabrol5 | Romain David1 | Emilie Egea1 | Jean-Baptiste Ledoux6,7 | Bastien Mérigot8 | Alexandra Anh-Thu Weber1,9 | Anne Chenuil1

1Aix Marseille Univ, Univ Avignon, CNRS, IRD, IMBE, Marseille, France
2Biology Department, Albion College, Albion, MI, USA
3Département de Biologie, Faculté des Sciences de la Nature et de la Vie, Université Abdelhamid Ibn Badis, Mostaganem, Algérie
4PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, Perpignan, France
5CNRS, Centrale Marseille, I2M, UMR7373, Aix-Marseille Université, Marseille, France
6CIMAR/CIMAR, Centro Interdisciplinar de Investigación Marina e Ambiental, Universidade do Porto, Porto, Portugal
7Institute of Marine Sciences (ICM-CSIC), Barcelona, Spain
8UMR MARBEC (CNRS, Ifremer, IRD, UM), Université de Montpellier, Sète, France
9Zoological Institute, University of Basel, Basel, Switzerland

Correspondence
Abigail E. Cahill, Biology Department, Albion College, Albion, MI, USA.
Email: acahill@albion.edu

Funding information
7th framework Programme, Grant/Award Number: 308392; Fundação para a Ciência e a Tecnologia, Grant/Award Number: SFRH/BPD/74400/2010; French National Research Agency, Grant/Award Number: ANR-11-LABX-0061. ANR-11-IDEX-0001-02. 12-SEAS-0001-01

Abstract
Genetic diversity is crucial for species’ maintenance and persistence, yet is often overlooked in conservation studies. Species diversity is more often reported due to practical constraints, but it is unknown if these measures of diversity are correlated. In marine invertebrates, adults are often sessile or sedentary and populations exchange genes via dispersal of gametes and larvae. Species with a larval period are expected to have more connected populations than those without larval dispersal. We assessed the relationship between measures of species and genetic diversity, and between dispersal ability and connectivity. We compiled data on genetic patterns and life history traits in nine species across five phyla. Sampling sites spanned 600 km in the northwest Mediterranean Sea and focused on a 50-km area near Marseilles, France. Comparative population genetic approaches yielded three main results. (i) Species without larvae showed higher levels of genetic structure than species with free-living larvae, but the role of larval type (lecithotrophic or planktotrophic) was negligible. (ii) A narrow area around Marseilles, subject to offshore advection, limited genetic connectivity in most species. (iii) We identified sites with significant positive contributions to overall genetic diversity across all species, corresponding with areas near low human population densities. In contrast, high levels of human activity corresponded with a negative contribution to overall genetic diversity. Genetic diversity within species was positively and significantly linearly related to local species diversity. Our study suggests that local contribution to overall genetic diversity should be taken into account for future conservation strategies.

Keywords
dispersal, genetic diversity, larvae, marine invertebrates, population structure, species genetic diversity correlation
Populations form the core of both ecological and evolutionary processes and are connected through the dispersal of organisms and genes (Palumbi, 2003). In the marine environment, many organisms display a benthic-pelagic lifestyle, with benthic adults producing planktonic larvae. It is this early life stage that frequently influences connectivity among marine populations (Cowen & Sponaugle, 2009).

However, directly measuring larval dispersal in the ocean is rarely possible due to the small propagule size and long dispersal distances of most species (but see Olson & McPherson, 1987). Instead, researchers rely on indirect measures of connectivity (López-Duarte et al., 2012), most commonly genetic differentiation among populations as measured by molecular markers.

In addition to hundreds of single-species analyses currently in the literature, another important line of research in marine connectivity has been compiling these studies to see what traits impact population differentiation. Investigators look at both closely related pairs of species that differ in key life history traits such as larval type (e.g., Cahill & Levinton, 2016; Cahill & Viard, 2014; Lee & Boulding, 2014) and larval dispersal potential (Shanks, Grantham, & Carr, 2003). Thus, PLD is expected to be negatively correlated with dispersal potential (Shanks, Grantham, & Carr, 2003). This, in turn, guides the development of effective conservation goals.

Such goals often include the conservation of biodiversity on all scales, including within species (genetic diversity) and among species (species diversity). As species diversity may be easier and less expensive to survey than genetic diversity, a positive correlation between these measures of diversity could lead to an effective policy to protect both the numbers of species and the evolutionary potential within those species. The higher the genetic diversity within a population or species, the lower the extinction risk (Reed & Frankham, 2003). Genetic diversity has also been proposed to impact function at higher levels of ecological organization (Barbour et al., 2016; Hughes, Inouye, Johnson, Underwood, & Vellend, 2008; Hughes & Stachowicz, 2004; Reusch, Ehlers, Hämmerli, & Worm, 2005) and is thus a major factor influencing a species’ resilience to perturbations. The patterns of within- and among-species diversity are not necessarily independent, because they may be generated by the same forces (e.g., patch size and distance from other patches; Vellend & Geber, 2005; Vellend, 2010; Lamy et al., 2013; Vellend et al., 2014).

One way to examine the relationship between dispersal and genetic structure is to analyse data from multiple species which co-occur in the same area, minimizing the confounding effects of geography (e.g., Kelly & Palumbi, 2010; Liggins, Treml, Possingham, & Riginos, 2016; Selkoe et al., 2014; Villamor et al., 2014). In the Mediterranean Sea, a recognized biodiversity hot spot (Bianchi & Morri, 2000), several single-species studies have investigated connectivity on regional and local scales (Table 1), and many have identified genetic structure even on a small scale. Barriers to gene flow have been identified between the eastern and western basins and sometimes within basins (Borsa et al., 1997; Chaouli, Kara, Quignard, Faure, & Bonhomme, 2009; Masmoudi et al., 2016; Pascual, Rives, Schunter, & Macpherson, 2017). The northwest basin is of particular interest in light of climate change as many species are already near their upper thermal limits and are unable to migrate further towards the poles (Lejeunesne, Chevaudonné, Per gent-Martini, Boudouresque, & Pérez, 2010). Its biota is already showing stress from climate change, including range shifts within the basin and large-scale mass mortality events linked to thermal anomalies (Francour, Boudouresque, Harmelin, Harmelin-Vivien, & Quignard, 1994; Garrabou et al., 2009; Lejeunesne et al., 2010; Marbà, Jordà, Agustí, Girard, & Duarte, 2015).

Within the northwest Mediterranean, an important habitat is the coralligenous reefs made by bioconstructors like coralline algae and gorgonians (Ballesteros, 2006; Gatti et al., 2012). These habitats are patchily distributed throughout the Mediterranean, with important concentrations around the Bay of Marseilles (Martin et al., 2014). This habitat is a priority for conservation as a nursery for fish and because it harbours a high level of species diversity (Ballesteros, 2006; Doxa et al., 2015; Gatti, Bianchi, Morri, Montefalcone, & Sar toretto, 2015). The northwest Mediterranean also contains protected areas (e.g., Parc National des Calanques, Parc National de Port Cros, Porquerolles, Natural Reserve of Scandola). The synthesis of previously identified genetic structure across several species would allow for the identification of particular barriers to gene flow that might be common to these species (DeBiasse, Richards, Shivy, & Hellberg, 2016; Lukoschek, Riginos, & van Oppen, 2016), in turn guiding conservation goals.
<table>
<thead>
<tr>
<th>Species</th>
<th>Abbrev.</th>
<th>Phylum</th>
<th>Data source</th>
<th>Number of pops</th>
<th>Loci used</th>
<th>N per pop (range)</th>
<th>Total N</th>
<th>Zones missing</th>
<th>Larval type</th>
<th>PLD (days)</th>
<th>Habitat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amphipholis squamata</td>
<td>asqu</td>
<td>Echinodermata</td>
<td>Boissin et al. (2008), Boissin, Egea, Féral, and Chenui (2015)</td>
<td>9</td>
<td>16S</td>
<td>7–30</td>
<td>154</td>
<td>5</td>
<td>Direct development</td>
<td>0</td>
<td>Coralline algae on rocky shores, eelgrass</td>
</tr>
<tr>
<td>Corallium rubrum</td>
<td>crub</td>
<td>Cnidaria</td>
<td>Ledoux et al. (2010), this study</td>
<td>39</td>
<td>10 msats</td>
<td>21–52</td>
<td>1,222</td>
<td>6</td>
<td>Lecithotrophic; internal fertilization and brooding</td>
<td>4–12</td>
<td>Coralligenous reefs and rocky habitats</td>
</tr>
<tr>
<td>Echinocardium cordatum</td>
<td>ecor</td>
<td>Echinodermata</td>
<td>Egea et al. (2016), this study</td>
<td>13</td>
<td>16S</td>
<td>6–43</td>
<td>248</td>
<td>3,7,8,9</td>
<td>Planktotrophic</td>
<td>15–40</td>
<td>Infaunal (sand)</td>
</tr>
<tr>
<td>Hemimysis margoleti</td>
<td>hmar</td>
<td>Arthropoda</td>
<td>Lejeusne and Chevaldonné (2006), Rastorgueff, Chevaldonné, Arslan, Verna, and Lejeusne (2014)</td>
<td>12</td>
<td>COI</td>
<td>5–18</td>
<td>106</td>
<td>1,7,9</td>
<td>Direct development</td>
<td>0</td>
<td>Marine caves</td>
</tr>
<tr>
<td>Myriopora truncata</td>
<td>mtru</td>
<td>Bryozoa</td>
<td>This study</td>
<td>16</td>
<td>COI</td>
<td>7–78</td>
<td>514</td>
<td>6</td>
<td>Lecithotrophic</td>
<td>1</td>
<td>Coralligenous reefs and shaded rocky shores</td>
</tr>
<tr>
<td>Ophioderma longicauda</td>
<td>olan</td>
<td>Echinodermata</td>
<td>Weber et al. (2015), this study</td>
<td>11</td>
<td>COI, 4 nuclear loci (1972, 55384, 79905, 147510)</td>
<td>7–32</td>
<td>249</td>
<td>Lecithotrophic</td>
<td>6–10</td>
<td>Coralligenous reefs; rocky habitats</td>
<td></td>
</tr>
<tr>
<td>Paracentrotus lividus</td>
<td>pliv</td>
<td>Echinodermata</td>
<td>Penant, Aurelle, Féral, and Chenui (2013), this study</td>
<td>14</td>
<td>COI</td>
<td>9–33</td>
<td>300</td>
<td>–</td>
<td>Planktotrophic</td>
<td>25–30</td>
<td>Mainly rocky shores and seagrass meadows</td>
</tr>
<tr>
<td>Paramuricea clavata</td>
<td>pcla</td>
<td>Cnidaria</td>
<td>Mokhtar-Jamai et al. (2011), this study</td>
<td>29</td>
<td>5 msats</td>
<td>28–62</td>
<td>952</td>
<td>–</td>
<td>Lecithotrophic; surface brooder</td>
<td>6–23</td>
<td>Coralligenous reefs</td>
</tr>
<tr>
<td>Patella caerulea</td>
<td>pcae</td>
<td>Mollusca</td>
<td>This study</td>
<td>10</td>
<td>COI</td>
<td>5–26</td>
<td>141</td>
<td>4,6,9</td>
<td>Lecithotrophic</td>
<td>7–10</td>
<td>Rocky shore</td>
</tr>
</tbody>
</table>
or negative (Vellend et al., 2014). Messmer, Jones, Munday, and Planes (2012) found concordance between genetic and species diversity patterns in tropical reef fish, although only three sites were investigated in this study. Similar to coral reefs, coralligenous reefs are patchily distributed (Martin et al., 2014), but are connected by larval dispersal of many species, so it is unclear if these communities would show strong SGDCs.

The goals of this study were (i) to test whether larval type explains patterns of connectivity at different scales; (ii) to identify the locations of genetic breaks at a small spatial scale (a few kilometres), particularly in the zone around the Bay of Marseilles where structure has been observed in multiple single-species studies; and (iii) to identify sites with high genetic diversity within the study zone and test the SGDC in this area. We synthesized population genetic data from nine species representing five phyla to assess connectivity in the northwest Mediterranean basin. The species represented a variety of larval development types, including two planktotrophs, five lecithotrophs and two direct developers (Table 1), and all were sampled across a wide geographic range within the study system. We compiled previously existing data and collected new data to enlarge and complete our data set. We expected that planktotrophs would be more connected across the study region than lecithotrophs and that direct developers would be the least connected, and that diversity would be positively correlated between the genetic and species levels.

2 | METHODS

2.1 | Study zone

Sampling sites in the northwest Mediterranean Sea ranged from Spain to Corsica, a coastline distance of approximately 600 km. Most sites were concentrated in a 50-km area near Marseilles, France (Table 2, Figure 1, Appendix S1). The prevailing current in this region is from east to west. Although wind forcing and coastal dynamics create complex patterns of currents in the area, particularly near the coast, all sampling sites fall into the same ecoregion (Berline, Ram mou, Doglioli, Molcard, & Petrenko, 2014). For the purpose of this study, we divided the basin a priori into nine sampling zones (Table 2, Figure 1). These zones were delineated visually based on the spacing of sampling sites as well as on patterns observed in previous studies (Table 1) and prior knowledge of the general ocean circulation patterns in the Bay of Marseilles and the Mediterranean Sea (Berline et al., 2014; Pradal & Millet, 2013), see Appendix S1 for all sampling sites for each species.

### TABLE 2

Geographic information about the sampling zones.

<table>
<thead>
<tr>
<th>Name</th>
<th>Number</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalonia</td>
<td>1</td>
<td>42°32’N</td>
<td>3°5’E</td>
</tr>
<tr>
<td>Côte Bleue</td>
<td>2</td>
<td>43°18’N</td>
<td>5°10’E</td>
</tr>
<tr>
<td>Marseilles</td>
<td>3</td>
<td>43°16’N</td>
<td>5°19’E</td>
</tr>
<tr>
<td>Calanques</td>
<td>4</td>
<td>43°10’N</td>
<td>5°24’E</td>
</tr>
<tr>
<td>Cassis</td>
<td>5</td>
<td>43°10’N</td>
<td>5°32’E</td>
</tr>
<tr>
<td>Porquerolles</td>
<td>6</td>
<td>42°58’N</td>
<td>6°13’E</td>
</tr>
<tr>
<td>St. Raphaël</td>
<td>7</td>
<td>43°24’N</td>
<td>6°46’E</td>
</tr>
<tr>
<td>Côte d’Azur</td>
<td>8</td>
<td>43°41’N</td>
<td>7°18’E</td>
</tr>
<tr>
<td>Corsica</td>
<td>9</td>
<td>42°23’N</td>
<td>8°34’E</td>
</tr>
</tbody>
</table>

### FIGURE 1

Map of the study area, divided into zones. Zone numbers correspond to those in Table 2. (a) Mediterranean Sea. (b) northwest Mediterranean. (c) Provence-Alpes-Côte d’Azur, France. (d) Bouches-du-Rhône, France. The dotted lines in (d) represent the borders of the zones, and the arrow represents Cape Croisette. See Appendix S1 for detailed sampling information for all sites and species.
2.2 | Data collection: existing data

Existing population genetic data sets representing seven species were compiled for this study (four echinoderms, two cnidarians and one arthropod; Table 1). The existing data sets included both mitochondrial and nuclear markers. The gorgonians *Corallium rubrum* and *Paramuricea clavata* were analysed with microsatellites because of a lack of variation in mitochondrial markers in these species (Calderón, Garrabou, & Aurelle, 2006); all other species were analysed using mitochondrial markers. Several populations (minimum = 9) were sampled for each species, with 5–78 individuals sampled per population. In a few cases (e.g., *Hemimysis margaefi*, *Echinocardium cordatum*), some populations in the original data sets contained fewer than five individuals and were removed from this study. In the cases of *Anphipholis squamata* and *E. cordatum*, the original data sets contained evidence of cryptic species (Boissin, Féral, & Chenuil, 2008; Egea et al., 2016). We retained individuals only from the most common clade in each species complex (clade A3 in *A. squamata* and B2 in *E. cordatum*). In the case of *P. clavata*, the original data set contained populations from the same site sampled at different depths. For this study, we chose a single population per site: the one sampled at a depth closest to 20 m, which corresponded to the sampling depth at other sites. See Appendix S1 for detailed information about the data sets used in this study.

2.3 | Data collection: new data

In five cases (*C. rubrum*, *E. cordatum*, *Ophioderma longicauda*, *P. clavata* and *Paracentrotus lividus*), we supplemented our data set with additional individuals, populations and markers, as well as adding two species (*Myriapora truncata* and *Patella caerulea*; Table 1). Details of collection and DNA extraction for each of these species, as well as sequencing for *M. truncata* and *P. clavata*, are listed in Appendices S1 and S2. For new samples of *E. cordatum*, *O. longicauda* and *P. lividus*, and all samples of *P. caerulea*, mtDNA markers were amplified individually. Primer sequences and PCR conditions are in Appendix S2. Four nuclear loci from *O. longicauda* were chosen and primers developed based on the transcriptomes for the species (Weber et al., 2015, 2017); the details of the methods and results for these markers are located in Appendix S3. All primers were labelled for use in illumina sequencing. Amplification success was determined visually using 1.5% agarose gels.

Following PCR amplification, the new samples were pooled for sequencing. Amplicons from multiple species were pooled in each well of three 96-well plates, with one well per plate left as a blank (285 wells total). Only one individual per species was added to each well. The volume added for each marker was determined based on its intensity on the agarose gels (classified as weak, medium or strong; see Appendix S2 for the exact volume of each marker added to the pool); 50 µl of each amplicon pool was purified using a NucleoFast® 96 PCR Clean-Up Kit (Machery-Nagel) according to the manufacturer’s instructions. After purification, 50-µl aliquots of the pools were sent for sequencing at Génopole Toulouse Midi-Pyrénées (GenoToul), where libraries were prepared for 2 × 250 bp paired-end sequencing on an Illumina Miseq. All 285 libraries were sequenced on a single standard Miseq run (15 million paired reads).

Because each individual well received a separate adapter for sequencing, individuals within species remained identifiable.

2.4 | Data analysis: bioinformatics

The sequencing run generated 10.9 million reads (average coverage per library of pooled amplicons = 19,000 reads, range per library 8,300 reads–75,600 reads). Following demultiplexing of forward and reverse reads, we filtered and trimmed all reads based on a mean Phred score of 35, calculated over a 10-bp sliding window. For each locus, we calculated the length to which we would subsequently trim all reads. The threshold for this second trimming was calculated as either the length X at which 95% of reads were of length X bp or longer, or as 200 bp, whichever was larger. Thresholds were chosen for forward and reverse reads separately. Reads were then truncated to the selected threshold, and forward and reverse reads were assembled. In several cases, the amplicons were so long that forward and reverse reads did not overlap. In these cases, we attached the forward and reverse reads end to end.

Following assembly, we removed singletons and counted the unique reads at each locus for each individual. We calculated the total coverage, as well as the ratio of the second most abundant read to the most abundant (count ratio). For mitochondrial loci, individuals were retained if the total number of reads was ≥10 (i.e., 10× coverage) and the count ratio was ≤0.3 (i.e., the most abundant read was at least 3 times abundant as the second most abundant read). Python scripts for our bioinformatics analyses are available at Github (https://github.com/chaby/dana/tree/master/src).

2.5 | Data analysis: within species

We calculated population genetic metrics for each species and marker in the data set. Previously collected data sets were reanalysed for this study, and in species for which we had old and new data at the same loci, the data were combined. In some cases (e.g., COI in *P. lividus*), this necessitated trimming the sequences in the older, Sanger-sequenced data set to match the read length in the Illumina data set.

For both mitochondrial and microsatellite markers, we used Genalex 6.5 (Peakall & Smouse, 2006, 2012) to calculate allelic richness or number of haplotypes (*Ar/Nh*). In *P. clavata* and *C. rubrum*, Genalex was also used to calculate overall *F*<sub>ST</sub>, pairwise *F*<sub>ST</sub>-based heterozygosity (*H*<sub>E</sub>) and Jost’s *D*<sub>est</sub>. For comparisons among data sets, the last two parameters were preferred over *Ar/Nh* and *F*<sub>ST</sub> because (i) *H*<sub>E</sub> varies between zero and one regardless of sample sizes, contrary to *Ar/Nh*, and (ii) *D*<sub>est</sub> is not affected by polymorphism contrary to *F*<sub>ST</sub>, whose maximum value is limited by *H*<sub>E</sub> (Jost, 2008). In all other species, we used ARLEQUIN 3.5 (Excoffier & Lischer, 2010) to calculate overall *F*<sub>ST</sub>, pairwise *F*<sub>ST</sub>, and *H*<sub>E</sub> based on complete haplotypes, as well as the significance of these metrics. The significance of pairwise *F*<sub>ST</sub> values was determined using
permuted tests in ARLEQUIN. \( D_{est} \) was calculated using SPADE (Chao & Shen, 2010). In all cases, we investigated isolation by distance with Mantel tests using GENEPOP on the Web (Rousset, 2008). Geographic distance was calculated as a straight-line distance from the nearest coastal site for islands, or the coastline distance between sampling sites along the coast (the straight-line distance between coastal sites often crossed land and the particular depth profiles that each species’ larvae use during development are largely unknown; distances are found in Appendix S4). Each analysis was conducted both using the whole data set (hereafter “global”) and using only those data from the area around Marseilles, where many species were sampled more densely (hereafter “local”). This corresponds to zones 2–4 (Table 2, Figure 1). We focused on this area for the local studies because previous single-species studies (Table 1) often identified genetic breaks on this small spatial scale.

2.6 | Analyses of multispecies genetic structure

We conducted analyses of variance to assess genetic differentiation according to larval type based on two measures of differentiation: Jost’s \( D_{est} \) and mean pairwise \( F_{ST} \). We focus on \( D_{est} \) rather than \( F_{ST} \) when comparing across species because \( F_{ST} \) is affected by marker type and polymorphism (Jost, 2008), and our data set combined both mitochondrial and microsatellite data. Analyses were conducted both on the global and local values of these metrics. We analysed a single mitochondrial locus for each species (except the two gorgonians, C. rubrum and P. clavata, for which we used the complete microsatellite data set). Analyses were performed in R 3.2.2 (R Core Development Team 2015).

To compare the geographic pattern of genetic structure across species at the global scale, we selected target sites for each species within each zone, such that the reduced data set for each species was comprised of data from \( >9 \) sites. When a species was sampled in more than one site per zone, we chose target sites based first on sample size (larger sample sizes were selected) and then according to the number of other species sampled at a site (sites with more species were selected). The target sites for each species are given in Appendix S1. The use of target sites gave us a more balanced design across sites and species, minimizing the amount of missing data in our analysis (Liggins et al., 2016).

To identify genetic breaks among the target sites, we used the function monmonier in the ADEGENET package, version 2.0.0 (Jombart, 2008) in R 3.2.2. The function uses Monmonier’s algorithm (Monmonier, 1973) to identify boundaries in genetic data based on genetic distances among populations and the geographic coordinates of those populations (Manni, Guerard, & Heyer, 2004), although it does not determine the causes of these boundaries. We identified the strongest boundary in each species using the default minimum genetic difference threshold between neighbours (0.1) and data from all nine target sites (i.e., the global scale). Genetic distances were calculated using the Euclidean distance between individuals, based on allelic data. The connection network was calculated using a Gabriel graph. When the default threshold difference of 0.1 did not identify any boundaries for a given species, we lowered the threshold until a boundary was found. As species had different overall levels of genetic structure, altering the threshold on a species-by-species basis was necessary to find the locations of strongest boundaries. For each pair of adjacent zones, we then calculated the proportion of species that had their strongest boundary between those zones, considering both the case where all species had a threshold of 0.1, and the case where the strongest boundary was identified, no matter the threshold value.

2.7 | Geographic patterns of genetic and species diversity

The relative contribution of each zone to the overall genetic diversity of the species was assessed using the program CONTRIB version 1.02 (Petit, El Mousadik, & Pons, 1998). This program calculates the contribution of each site to overall genetic diversity based on expected heterozygosity. This contribution \( C_s \) is the sum of two components, one \( (C_{sP}) \) reflecting each population’s divergence from other populations and one \( (C_G) \) which is due to its level of diversity. For each species, we calculated the \( C_S \) value for each of the nine target sites (i.e., the global scale). Positive values of \( C_S \) indicate a positive contribution of a zone to overall diversity, and negative values indicate a negative contribution (Petit et al., 1998). We took the mean \( C_S \) value for each zone and evaluated its difference from zero using 95% confidence intervals.

We analysed the relationship (slope and correlation) between genetic and species diversity using species diversity data from the observer network of coralligenous habitats (Réseau du suivi des assemblages coralligènes, REC2R; Andromede Oceanologie 2013). These data were collected using analyses of photoquadrats in coralligenous reefs (described in Deter, Descamp, Boisserie, Ballesta, & Holon, 2012; Doxa et al., 2015). Sampling was conducted from 2010 to 2014, at depths of 35–63 m (Appendix S5). At each station, 30 quadrats (0.5 m²) were photographed along a 40-m transect (Doxa et al., 2015). For our analyses, we selected one representative station from the REC2R database in each zone (Appendix S5). We selected the stations with available species diversity data that were as close as possible to the nine target sites for the population genetic data. We conducted a correlation analysis between the Simpson’s index for species diversity at these stations (provided by REC2R at http://www.observatoire-mer.fr/) and expected heterozygosity for genetic diversity \( (H_E) \) because their mathematical formula is the same. When multiple sampling dates or multiple depths were available at a given site, we averaged the Simpson’s index values to get a single-species diversity value for the site. We replaced missing genetic diversity values at a given site with the mean genetic diversity for the species.

3 | RESULTS

3.1 | Genetic diversity metrics across each species

Values of gene diversity \( (H_E) \) in the global data set ranged from 0.003 (E. cordatum) to 0.787 (P. clavata; Table 3). In the local data
set, expected heterozygosity values ranged from 0.0003 (E. cordatum) to 0.781 (P. clavata; Table 4). FST values in the global data set ranged from 0.003 (E. cordatum) to 0.621 (M. truncata; Table 3). In the local data set, FST values ranged from −0.007 (O. longicauda) to 0.516 (A. squamata; Table 4). Values of Dst ranged from −0.03 (E. cordatum) to 0.807 (M. truncata) in the global data set and from −0.094 (E. cordatum) to 0.969 (A. squamata) in the local data set (Tables 3 and 4).

### 3.2 Direct developers display more genetic structure than species with planktonic larvae

Neither mean Dst nor mean pairwise FST was significantly different across larval types in the global data set (DST: F2,6 = 2.909, p = .131; pairwise FST: F2,6 = 1.223, p = .359; Figure 2a,c). However, when M. truncata was removed from the data set (see Discussion), both analyses became significant (DST: F2,5 = 14.02, p = .009; pairwise FST: F2,5 = 36.61, p = .001), with direct developing species having significantly more structure than species with either planktotrophic or lecithotrophic larvae according to Tukey’s HSD tests. Within the local data set (all species included), both metrics were significantly different across larval types (DST: F2,6 = 16.93, p = .003; pairwise FST: F2,6 = 11.73, p = .008). In both cases, Tukey’s HSD tests showed that direct developing species had significantly more structure than planktotrophic and lecithotrophic species (Figure 2b,d).

### 3.3 Spatial patterns of genetic diversity

Significant positive patterns of isolation by distance were observed in the global data set in three species: C. rubrum (lecithotroph), H. margoleti (direct developer) and P. clavata (lecithotroph; Table 3). All three species also showed significant isolation by distance in the local data set (Table 4). Full pairwise FST matrices and IBD graphs for each species are in Appendix S6.

The CONTRIB analysis showed that the mean Cs values across all species were highest in Corsica (Zone 9; mean = 0.027 ± 0.011 SE) and lowest in the Côte Bleue (Zone 2; mean = −0.028 ± 0.009 SE); Cs was also significantly positive in Porquerolles (Zone 6; mean = 0.013 ± 0.006 SE). Cs values for all other sites were not significantly different from zero (Figure 3). A t test of Cs values comparing the two insular zones (6 and 9) to all others was significant (t = −3.171, p = .004). Complete results of the CONTRIB analysis for each species are in Appendix S6.

Monmonier’s algorithm found boundaries in five species with a difference threshold of 0.1 (Figure 4, Appendix S7). At this level of differentiation, no boundaries were found in the global area for P. caerulea, O. longicauda, P. lividus or E. cordatum, although boundaries between populations were found in these species by lowering the threshold of differentiation to 0.01, 0.0025, 0.005 or 0.0025, respectively (Figure 4). When only boundaries found at the threshold of 0.1 were considered, the greatest proportion of them were found

### TABLE 3 Global genetic diversity metrics of mitochondrial and microsatellite loci. Values are underlined if p < .05 (value is significantly different from zero). IBD slopes are in underline if p < .05 according to a Mantel test (slope is significantly different from zero) and if the slope is positive. All sites sampled were included in these analyses. Species abbreviations correspond to those in Table 1, and species are ordered by increasing PLD

<table>
<thead>
<tr>
<th>Species</th>
<th>Development type</th>
<th>Marker</th>
<th>FST</th>
<th>Dst</th>
<th>Ht</th>
<th>IBD slope</th>
<th>Mean pairwise FST</th>
</tr>
</thead>
<tbody>
<tr>
<td>asqu</td>
<td>Direct</td>
<td>16S</td>
<td>0.479</td>
<td>0.578</td>
<td>0.064</td>
<td>−4.16 × 10⁻³</td>
<td>0.380</td>
</tr>
<tr>
<td>hmar</td>
<td>Direct</td>
<td>COI</td>
<td>0.244</td>
<td>0.695</td>
<td>0.136</td>
<td>1.52 × 10⁻³</td>
<td>0.263</td>
</tr>
<tr>
<td>mtru</td>
<td>Lecithotroph</td>
<td>COI</td>
<td>0.621</td>
<td>0.807</td>
<td>0.029</td>
<td>−2.44</td>
<td>0.581</td>
</tr>
<tr>
<td>crub</td>
<td>Lecithotroph</td>
<td>10 msats</td>
<td>0.119</td>
<td>0.304</td>
<td>0.7200</td>
<td>1.57 × 10⁻⁵</td>
<td>0.064</td>
</tr>
<tr>
<td>pcla</td>
<td>Lecithotroph</td>
<td>5 msats</td>
<td>0.009</td>
<td>0.293</td>
<td>0.787</td>
<td>2.2 × 10⁻⁵</td>
<td>0.044</td>
</tr>
<tr>
<td>pcae</td>
<td>Lecithotroph</td>
<td>COI</td>
<td>0.199</td>
<td>0.032</td>
<td>0.025</td>
<td>8.95 × 10⁻⁶</td>
<td>0.004</td>
</tr>
<tr>
<td>olon</td>
<td>Lecithotroph</td>
<td>COI</td>
<td>0.046</td>
<td>0.044</td>
<td>0.005</td>
<td>5.32 × 10⁻⁶</td>
<td>0.035</td>
</tr>
<tr>
<td>ecor</td>
<td>Planktotroph</td>
<td>16S</td>
<td>0.003</td>
<td>−0.030</td>
<td>0.004</td>
<td>−3.20 × 10⁻⁵</td>
<td>−0.005</td>
</tr>
<tr>
<td>pliv</td>
<td>Planktotroph</td>
<td>COI</td>
<td>0.019</td>
<td>0.053</td>
<td>0.006</td>
<td>1.61 × 10⁻⁵</td>
<td>0.018</td>
</tr>
</tbody>
</table>

IBD, isolation-by-distance; PLD, planktonic larval duration.

### TABLE 4 Genetic diversity metrics of mitochondrial and microsatellite loci in the Marseilles area (zones 2-4). Values are underlined if p < .05 (value is significantly different from zero). IBD slopes are in underline if p < .05 according to a Mantel test (slope is significantly different from zero) and if the slope is positive. All sampled sites within zones 2, 3 and 4 were included in these analyses. Species abbreviations correspond to those in Table 1, and species are ordered by increasing PLD

<table>
<thead>
<tr>
<th>Species</th>
<th>Development type</th>
<th>Marker</th>
<th>FST</th>
<th>Dst</th>
<th>Ht</th>
<th>IBD slope</th>
<th>Mean pairwise FST</th>
</tr>
</thead>
<tbody>
<tr>
<td>asqu</td>
<td>Direct</td>
<td>16S</td>
<td>0.516</td>
<td>0.969</td>
<td>0.117</td>
<td>2.11 × 10⁻¹</td>
<td>0.530</td>
</tr>
<tr>
<td>hmar</td>
<td>Direct</td>
<td>COI</td>
<td>0.216</td>
<td>0.565</td>
<td>0.190</td>
<td>2.12 × 10⁻²</td>
<td>0.215</td>
</tr>
<tr>
<td>mtru</td>
<td>Lecithotroph</td>
<td>COI</td>
<td>0.023</td>
<td>0.012</td>
<td>0.001</td>
<td>−2.00 × 10⁻⁴</td>
<td>0.023</td>
</tr>
<tr>
<td>crub</td>
<td>Lecithotroph</td>
<td>10 msats</td>
<td>0.105</td>
<td>0.260</td>
<td>0.722</td>
<td>4.0 × 10⁻⁴</td>
<td>0.057</td>
</tr>
<tr>
<td>pcla</td>
<td>Lecithotroph</td>
<td>5 msats</td>
<td>0.055</td>
<td>0.185</td>
<td>0.781</td>
<td>9.7 × 10⁻⁴</td>
<td>0.032</td>
</tr>
<tr>
<td>pcae</td>
<td>Lecithotroph</td>
<td>COI</td>
<td>0.014</td>
<td>0.028</td>
<td>0.032</td>
<td>−1.24 × 10⁻³</td>
<td>0.012</td>
</tr>
<tr>
<td>olon</td>
<td>Lecithotroph</td>
<td>COI</td>
<td>−0.004</td>
<td>−0.009</td>
<td>0.003</td>
<td>2.51 × 10⁻³</td>
<td>−0.003</td>
</tr>
<tr>
<td>ecor</td>
<td>Planktotroph</td>
<td>16S</td>
<td>−0.042</td>
<td>−0.094</td>
<td>0.0003</td>
<td>−</td>
<td>−0.042</td>
</tr>
<tr>
<td>pliv</td>
<td>Planktotroph</td>
<td>COI</td>
<td>0.001</td>
<td>−0.008</td>
<td>0.006</td>
<td>4.75 × 10⁻⁴</td>
<td>0.003</td>
</tr>
</tbody>
</table>

IBD, isolation-by-distance; PLD, planktonic larval duration.
between zones 1 and 2 and zones 2 and 3 (37.5% of species sampled in adjacent zones showed a boundary here). When all thresholds were considered, the greatest proportion of boundaries was found between zones 1 and 2 (75% of species), followed by zones 2 and 3 (50% of species).

Overall, species and genetic diversity were positively related \( (\gamma = 0.345x - 0.110, p = .02) \), and fitted values of genetic diversity varied only between 0.16 and 0.20 (Figure 5). The Pearson correlation coefficient was relatively high \( (r = .749; p = .02, \text{Figure 5}) \). No individual species showed a significant SGDC (Appendix S8), but in eight of nine species, the correlation was positive. The highest correlation coefficient (0.61) was observed in *P. clavata*.

**DISCUSSION**

**4.1 Development type determines genetic structure at regional and local scales**

Genetic structure was shown to be associated with development type at both large and small spatial scales, and using multiple metrics. At both spatial scales, genetic structure was associated with having a larva or not, rather than the type of larva (planktotroph vs. lecithotroph). This dichotomy between species with and without free-living larvae has been shown in the literature, both in analyses of large data sets (e.g., Haye et al., 2014; Kelly & Palumbi, 2010; Riginos et al., 2011; Selkoe & Toonen, 2011; Selkoe et al., 2014; Villamor et al., 2014; Weersing & Toonen, 2009; Wright et al., 2015) and in comparisons of closely related species (e.g., Boissin, Stöhr, & Chenuil, 2011; Cahill & Viard, 2014; Weber et al., 2015).

The fact that genetic structure was not different among larval types at the global scale was due to high levels of structure in the bryozoan *M. truncata* relative to other lecithotrophs. *M. truncata* has a very short larval duration (\( \leq 24 \) hrs, Ferretti, Magnino, & Balduzzi, 2007), which does not correspond to other species in the data set; all other lecithotrophs have longer larval stages, and indeed, the genetic structure in these species more closely resembles that of the planktotrophs.

The relationship between geographic distance and genetic structure was complex and varied among species. Some species showed a pattern of isolation by distance, and these species were either direct...
developers or lecithotrophs. However, some species of direct developers and lecithotrophs did not show IBD. The ability to recover IBD is related to both the scale of dispersal of a species and the density of sampling (Aurelle & Ledoux, 2013); no planktrophs showed IBD on the scale of this study.

4.2 Genetic breaks correspond to oceanographic currents in the Bay of Marseilles

It is noteworthy that species with larval periods longer than approximately 10 days did not show any strong boundaries when a threshold difference of 0.1 was used in the Monmonier algorithm. Many species showed strong genetic breaks between the Côte Bleue and the Catalan coast (zones 1 and 2). This is probably best explained by the distance between the sampling zones (approximately 250 km, the largest distance between adjacent zones in the study area), which may introduce false gaps into our results. However, there is also a lack of suitable habitat for organisms dependent on coralligenous reefs or other hard substrate, a biologically relevant reason for these breaks. The shallow habitat in the Gulf of Lions between zones 1 and 2 is dominated by sand and seagrass beds (Andromede Oceanologie 2013), which is not the habitat of most of the species in this study (except E. cordatum).

Most species showed significant structure in the Bay of Marseilles, with a high proportion of species showing their strongest boundary between zones 2 and 3. This structure is particularly striking given that the centres of these zones are only 12 km apart. Such a high level of structure on a small geographic scale is theoretically unexpected, and best explained by oceanographic currents. The currents in the Bay of Marseilles (zones 2–4) are influenced by two dominant wind regimes, with winds coming from either the northwest or from the southeast (Pradal & Millet, 2013). Under the northwestern wind regime, water masses move from zone 2 towards zone 3, but are rapidly driven offshore at Cape Croisette, the breakpoint between zones 3 and 4 (Pradal & Millet, 2013). It is highly likely that larvae with weak swimming abilities would be advected offshore, resulting in low connections in the Bay of Marseilles under this wind regime, which occurs 38% of the time (Pradal & Millet, 2013). A similar disconnect between these two zones has been found in the brown alga Cystoseira amentacea (Thibaut et al., 2016).

Circulation is more complicated under the southeast wind regime (occurs 15% of the time; Pradal & Millet, 2013), with water moving from east to west across Cape Croisette. This allows for more connectivity between zones 3 and 4, and indeed, most species do not show their strongest boundary in this area. More fine-scale oceanographic models are necessary to explore the relationship between timing of reproduction, larval duration, current patterns, wind frequencies and geography (Thibaut et al., 2016). Linking oceanographic patterns and genetic structure can help us to better understand patterns of connectivity and dispersal in the marine environment (Crandall, Trembl, & Barber, 2012; Liggins et al., 2016; Schiavina, Marino, Zane, & Melià, 2014; Selkoe et al., 2016). Small-scale models can provide insights which get lost in basin-scale models. Additionally, zone 2 is heavily impacted by commercial shipping and other human activities, and the genetic break between zones 2 and 3 may be due in part to these impacts. Small-scale models can therefore provide a baseline expectation against which to test the impact of human activity or other stressors.

4.3 Higher contribution to genetic diversity in island protected areas of low human population density

The CONTRIB analyses revealed high contributions to genetic diversity across all species in Corsica (zone 9) and Porquerolles (zone 6). Both zones are on islands (though of very different sizes and levels of isolation). As we used the C_S component of the diversity contribution, which reflects the diversity rather than the divergence of a population, the significantly positive contributions found in the two islands most likely reflect higher local effective population sizes and not their genetic isolation. Isolation from sources of migrants should lead to decreased genetic diversity locally, everything else being equal. Thus, the high C_S values of both islands strongly suggest that these samples are indeed subject to fewer perturbations than other sites, and suggest the effectiveness of marine protected areas as reservoirs of genetic diversity.
Disturbance is expected to have a negative impact on genetic diversity, based both on theory (Chenuil, 2012; Rossberg et al., 2017) and on empirical evidence, including from other marine systems (Ozerov, Veselov, Lumme, & Primmer, 2013; Pini, Planes, Rochel, Lecchini, & Fauvelot, 2011; Pinsky & Palumbi, 2014; Taris, Ernande, McCombie, & Boudry, 2006). Our data are consistent with the idea that anthropogenic impacts may influence overall diversity levels. In particular, Corsica is subjected to fewer anthropogenic impacts than mainland France. Additionally, both of these sampling zones are in or near national parks (Parc National de Port Cros-Porquerolles, Scandola Nature Reserve) and have low population densities (16 inhabitants/km² on Porquerolles and 21.27 inhabitants/km² on Corsica; data from the French National Institute of Statistics and Economics Studies). In contrast, the Côte Bleue (zone 2) showed consistently low contributions to genetic diversity across all species. This area is known to be heavily impacted by human activity, as Marseilles is the largest commercial port in France. The contributions to genetic diversity in all other zones were not different from zero when all species were considered. Zone 3 is also located in Marseilles but does not display significant negative contribution across species. However, most species in zone 3 were sampled in the Frioul archipelago, where the influence of mainland coastal activity is lower, and the two species sampled on the mainland coast actually display the lowest contributions to overall diversity for this zone. As for zone 4, also near Marseilles, local genetic diversity contributions are negative for all species except two.

**FIGURE 4** Strongest boundaries for each of nine species in zones 2–5. Map of the target sites in zones 2–5 showing the sampling zones (numbers) and the location of the strongest boundaries for each species, as identified using Monmonier’s algorithm. Target sites are placed on the coastline and solid lines between zone numbers indicate boundaries identified with a threshold of 0.1, while dashed lines between zone numbers indicate barriers identified with lowered thresholds (details in text). Colours indicate different species, arranged in the order of increasing planktonic larval duration. Triangles: direct developers; circles: lecithotrophs; squares: planktotrophs. Solid arrow indicates Cape Croisette, the boundary between zones 3 and 4. Information for all zones is in Appendix S7. This map illustrates that (i) zones 2 and 3 are much less connected genetically than zones 3 and 4, and zones 4 and 5 and (ii) boundaries are stronger and more numerous in species with lower dispersal ability.

**FIGURE 5** Relationship between species and genetic diversities. Relationship between the mean genetic diversity ($H_E$, expected heterozygosity, calculated across all species) and species diversity (Simpson’s index) at target sites within all nine zones. Pearson’s correlation $r = .749; (p = .02)$. Numbers correspond to zones. Missing values for each species were replaced with the mean genetic diversity for that species. Solid line indicates the best-fit line to the data ($y = 0.345x - 0.110, p = .02$). See Appendix S1 for a list of target sites in each species, Appendix S5 for information about the species diversity data and site list, and Appendix S8 for separate correlations for each species.
4.4 Genetic diversity is significantly correlated with species diversity

There was a significant linear relationship between total species diversity and average within-species genetic diversity in the study zones, although no SGDC was significant for any individual species. It should be noted our analysis had several missing values, which were replaced with species means. This significant correlation was found despite the fact that the genetic and species diversity data were collected using different methods and at different scales. The species diversity data come from a study based on photographs (Andromede Oceanologie 2013; Doxa et al., 2015) which may have underestimated diversity, as some species cannot be distinguished using this method. Sampling sites, dates and schemes did not overlap precisely between the two data types. We also see this correlation despite the fact that the sites that we compared are spaced close together in a matrix of ocean that allows dispersal of the target species; it is expected that open habitats will show weaker correlations between levels of diversity than isolated, closed habitats (Vellend et al., 2014). Messmer et al. (2012) found an association between diversity levels in coral reef fish in three sites of an analogous, open marine habitat. We suggest that future work investigates the SGDCs in coralligenous reef habitat using a more stringent sampling design and less missing data, coupled with techniques that are able to identify cryptic species, to confirm that species diversity may be used to predict genetic diversity.

4.5 Applications of the study in the design of marine protected areas

Patterns of connectivity in marine species, particularly comparative studies across species with different life history traits, can be used to help determine placement of and connections among marine protected areas (Drew & Barber, 2012; Eastwood, Lopez, & Drew, 2016; von der Heyden et al., 2014; Palumbi, 2003; Selkoe et al., 2016; Shanks et al., 2003). Our results indicated surprisingly complex patterns of connectivity and high genetic structure for some species, particularly within the Bay of Marseilles. Small-scale structure identified by fine-scale genetic sampling and fine-scale oceanographic mapping should be used when constructing networks of protected areas in this region.

Such networks should also take genetic diversity into account, aiming to maximize genetic diversity within protected areas (Munguía-Vega et al., 2015; Nielsen, Berger, Henriques, Selkoe, & von der Heyden, 2017; Pope, Riginos, Ovenden, Keyse, & Blomberg, 2015) as well as considering the connections between them (e.g., Beger et al., 2014). Our results show a high contribution to genetic diversity from two protected areas (Porquerolles and Corsica), although this was not observed in the Calanques zone, another protected area. The Calanques National Park (created in 2012) is much younger than both Porquerolles (Port-Cros National Park, 1963) and Corsica (Natural Reserve of Scandola, 1975). Another possible explanation for this discrepancy among protected areas is the proximity of the coastal Calanques zone to Marseilles as compared to either Porquerolles or Scandola, which are both located on islands. The use of multiple species in this study allowed us to identify areas where genetic diversity is higher overall, patterns that are not visible when considering each species individually. As patterns of genetic diversity and structure are usually different among species, such a multispecies approach is valuable to identifying high-priority areas for protection (e.g., Nielsen et al., 2017). Areas of high genetic diversity in this study were correlated with high species diversity, indicating that in this system and with these methods, it may be possible to predict genetic diversity hot spots based on species diversity hot spots or vice versa.

5 CONCLUSIONS

A synthesis of population genetic data from nine species sampled over the same area in the northeastern Mediterranean Sea allowed for a comparative analysis of genetic structure and diversity at these sites. As predicted, we found that genetic structure was stronger in species with low dispersal capacity (direct developers) than high dispersal capacity (species with free-living larvae) and that this difference was consistent on both large and small spatial scales. We found the Bay of Marseilles to have a surprising amount of genetic structure, given its small spatial scale, most likely due to local oceanographic currents. Areas of high genetic diversity corresponded to protected island areas and to high species diversity; these areas were identified by the use of multispecies data, which would not have been possible using single-species data. These results reiterate the importance of species traits and physical connectivity in generating population genetic structure, but also highlight the potential role of protected areas in maintaining genetic diversity.

ACKNOWLEDGEMENTS

Funding was provided by DEVOTES (DEvelopment Of innovative Tools for understanding marine biodiversity and assessing good Environmental Status), funded by the European Union under the 7th FRAMEWORK Programme, “The Ocean of Tomorrow” (grant agreement no. 308392), www.devotes-project.eu, and by La Faculté des Sciences de la Nature et de la Vie de l’Université Abdelhamid Ibn Badis, Mostaganem, Algeria (ZB) and the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia; FCT; Postdoctoral grant SFRH/BPD/74400/2010; JBL). This work is a contribution to Labex OT-Med (n° ANR-11-LABX-0061) and has received funding from Excellence Initiative of Aix-Marseille University - A*MIDEX, a French “Investissements d’Avenir” programme. Bryozoan samples were obtained thanks to EraNet (SeasEra) CIGESMED—ANR convention no 12-SEAS-0001-01. We thank GénéToul (particularly S. Valière) for sequencing, the Service Commun de Biologie Moléculaire of the IMBE at the Station Marine d’Endoume (A. Hagenauer, C. Ribout, C. Rocher, M. Selva), the Service Plongée of the Pythéas Institute, Port-Cros National Park for
help with sample collection and preparation, B. Millet for help regarding the oceanography in the Bay of Marseilles, E. Rollinson for help with figure preparation, and three anonymous reviewers for their helpful comments.

DATA ACCESSIBILITY

New sequences generated for this project are available in GenBank (Accession nos MG063882–MG064453). The raw data files from the Illumina sequencing run have been deposited at DataDryad (https://doi.org/10.5061/dryad.dv4mg), as well as new microsatellite genotypes and alignments for all loci that include haplotype frequency and population information. Details of accessibility are in Appendix S9. Bioinformatics code for processing the Illumina data is available on GitHub (https://github.com/chaby/dana).

AUTHOR CONTRIBUTIONS

A.C. conceived the study. The bioinformatics pipeline was developed by O.C. Data analyses and writing of the manuscript were carried out by A.E.C. All authors collected data and contributed to manuscript editing and revisions.

ORCID

Abigail E. Cahill http://orcid.org/0000-0002-3984-7212
Emilie Boissin http://orcid.org/0000-0002-4110-790X
Alexandra Anh-Thu Weber http://orcid.org/0000-0002-7980-388X
Anne Chenuil http://orcid.org/0000-0001-8141-7147

REFERENCES


**SUPPORTING INFORMATION**

Additional Supporting Information may be found online in the supporting information tab for this article.